

**Tuesday 17 July 2018****9h30****UT3 Paul Sabatier, IRIT, Salle des Thèses****Mahdi WASHHA****Team SIG - IRIT****Information Quality in Online Social Media and
Big Data Collection: An Example of Twitter Spam
Detection***Jury:*Pr. Anne BOYER – *Université de Lorraine, Rapporteur*Pr. Arnaud MARTIN – *Université de Rennes 1, Rapporteur*Pr. Morgan MAGNIN – *Laboratoire des Sciences de Nantes, Examineur*Pr. Josiane MOTHE – *Université Toulouse Jean Jaurès, Examineur*Pr. Florence SEDES – *Université Paul Sabatier, Directeur de Thèse*

Abstract: The popularity of Online Social Media (OSM) is mainly conditioned by the integrity and the quality of user-generated content (UGC) as well as the protection of users' privacy. Based on the definition of information quality as fitness for use, the high usability and accessibility of OSM have exposed many information quality (IQ) problems which consequently decrease the performance of OSM dependent applications. Such problems are caused by ill-intentioned individuals who misuse OSM services to spread different kinds of "noisy" information, including fake information, illegal commercial content, drug sales, malware downloads, and phishing links. The propagation and spreading of noisy information causes enormous drawbacks related to resources consumptions, decreasing quality of service of OSM-based applications, and wasting human efforts. The majority of popular social networks (e.g., Facebook, Twitter, etc) over the Web 2.0 is daily attacked by an enormous number of ill-intentioned users. However, those popular social networks are ineffective in handling such a "noisy" information, requiring several weeks or months to detect them. Moreover, different challenges stand in front of building a complete OSM-based information filtering methods that can overcome the shortcomings of OSM information filters. These challenges are summarized in: (i) big data; (ii) privacy and security; (iii) structure heterogeneity; (iv) UGC format diversity; (v) subjectivity and objectivity; (vi) and service limitations.

In this thesis, we focus on increasing the quality of social UGC that are published and publicly accessible in forms of posts and profiles over Online Social Networks (OSNs) through addressing in-depth the stated above serious challenges. As the social spam is the most common IQ problem appearing over the OSM, we introduce a design of two generic approaches for detecting and filtering out the spam content. The first approach is for detecting the spam posts (e.g., spam tweets) in a real-time stream, while the other approach is dedicated for handling a big data collection of social profiles (e.g., Twitter accounts). For filtering the spam content in real-time, we introduce an unsupervised collective-based framework that automatically adapts a supervised spam tweet classification function in order to have an updated real-time classifier without requiring manual annotated datasets. In the second approach, we treat the big data collections through minimizing the search space of profiles that needs advanced analysis, instead of processing every user's profile existing in the collections. Then, each profile falling in the reduced search space



is further analyzed in an advanced way to produce an accurate decision using a binary classification model.

The experiments conducted on Twitter online social network have shown that the unsupervised collective-based framework is able to produce updated and effective real-time binary tweet-based classification function that adapts the high evolution of social spammer's strategies on Twitter, outperforming the performance of two existing real-time spam detection methods. On the other hand, the results of the second approach have demonstrated that performing a preprocessing step for extracting spammy meta-data values and leveraging them in the retrieval process is a feasible solution for handling a large collection of Twitter profiles, as an alternative solution for processing all profiles existing in the input data collection...



05 61 55 65 10

info@irit.fr

www.irit.fr

